

Research paper

## Efficient dose-finding for drug combination studies involving a shift in study populations

Nolan A. Wages<sup>a,\*</sup>, Trish A. Millard<sup>b</sup>, Patrick M. Dillon<sup>b</sup>, Christiana M. Brenin<sup>b</sup>, Gina R. Petroni<sup>a</sup>

<sup>a</sup> Division of Translational Research & Applied Statistics, Department of Public Health Sciences, University of Virginia, Charlottesville, VA, USA

<sup>b</sup> Division of Hematology/Oncology, Department of Medicine, University of Virginia, Charlottesville, VA, USA



## ARTICLE INFO

## Keywords:

Dose finding  
Partial order continual reassessment method  
Drug combinations  
Patient heterogeneity  
Early-phase

## ABSTRACT

This paper describes the design of an early phase, prospective trial evaluating the safety and tolerability of the combination of the histone deacetylase inhibitor, entinostat, in combination with capecitabine. The study consists of two parts; an initial phase evaluating the safety of the combination in participants with metastatic breast cancer, followed by a second phase assessing the safety of the combination in participants with residual disease after neo-adjuvant chemotherapy for breast cancer. We describe the adaptation of a model-based design for identifying the maximum tolerated dose combination that efficiently moves from the initial phase in an advanced disease population to the second phase in the target population. Operating characteristics demonstrate the ability of the method to accurately predict true maximum tolerated dose combinations in a high percentage of trials with reasonable sample sizes, while treating participants at and around desirable combinations. The proposed design is a practical, early-phase, adaptive method for use with drug combination dose finding in the presence of shifting patient populations. More challenging research questions are being investigated in early-phase trials, which has created the need to implement more flexible designs that can meet the objectives of current studies, such as those exploring drug combinations while addressing patient heterogeneity. Our goal is to facilitate acceptance and application of more novel designs in contemporary early-phase studies.

### 1. Introduction

This article describes the design of an early phase, prospective trial evaluating the safety and tolerability of entinostat (a histone deacetylase inhibitor) in combination with capecitabine (a cytotoxic antimetabolite) in breast cancer participants (Breast 49, NCT03473639), designed at the University of Virginia (UVA) Cancer Center. Originally, the study design involved the application of a Bayesian continual reassessment method for drug combinations [1] in order to locate the maximum tolerated dose combination (MTDC) of the two drugs in participants with high risk breast cancer after neo-adjuvant chemotherapy. Review of the study protocol by the Food and Drug Administration (FDA) resulted in a request to “conduct a phase I dose escalation study in the metastatic setting. Once the safe dose for the combination is found, then the combination can be evaluated in the early breast cancer setting.” This request created the need to adapt the trial design in order to accommodate a change in patient populations. The trial objective shifted to

one of defining a safe dose combination in participants with metastatic breast cancer (Population A) and subsequently using the accumulated data from the metastatic group to inform the dose finding of the MTDC in participants with residual disease (Population B). The objective of dose escalation in Population A is to obtain an initial safety signal by evaluating the tolerability of low dose combinations. The goal is not to recommend an MTDC in Population A for further efficacy testing, but accrual of Population A participants is necessary in establishing a starting point for the accrual of Population B participants. It is expected that participants in Population B will better tolerate the combination than participants in Population A, so that if the combination is deemed safe for Population A, it would be inefficient to try non-toxic combinations on Population B. This provides justification for starting Population B at the estimated MTDC of Population A. The design described in this paper adapts the statistical modeling framework known as the partial order continual reassessment method (POCRM [1]) in order to incorporate potentially heterogeneous patient populations in the estimation

\* Corresponding author. Public Health Sciences, P.O. Box 800717 Charlottesville, Virginia, USA.

E-mail address: [nwages@virginia.edu](mailto:nwages@virginia.edu) (N.A. Wages).

<https://doi.org/10.1016/j.conctc.2020.100519>

Received 12 July 2019; Received in revised form 20 December 2019; Accepted 4 January 2020

Available online 7 January 2020

2451-8654/© 2020 The Authors.

Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

of the MTDC of the two agents being investigated. While the estimation of dose-limiting toxicity (DLT) probabilities utilizes POCRM throughout both phases of the study, execution of the design required necessary adaptations resulting from studying different patient populations. Adjustments to software code had to be made in order to evaluate specific aspects of this trial, such as simulating hypothesized DLT probability scenarios with different probabilities for each population, as well as the percentage of simulated trials in which the MTDC selected at the conclusion of the study was different than the initial MTDC estimated based on Population A. Consequently, specific aspects of this trial make a direct application of POCRM infeasible. This trial adds to a growing number of studies that have adapted POCRM to address specific research objectives presented by modern early-phase trials [2–4].

The advantage of the approach described herein over a traditional rule-based method lies in the ability of our model-based method to adjust to potential changes in the toxicity profile of the combinations when shifting study populations. Our design will sequentially update model-based toxicity probabilities based on all accumulated data and refine the recommended dose for subsequent cohorts. This approach makes efficient use of accumulated data in directing escalation and de-escalation decisions, potentially allowing fewer study participants to be treated at sub-therapeutic combinations. Additionally, the seamless nature of the design avoids the need for study protocol amendments that require adjustments to the initially identified MTDC from Population A. This approach is consistent with published recommendations on dose expansion cohorts that advocate for designs that can efficiently refine an initial MTD estimate from a different patient population [5,6]. Details of study designs often are not found on sites such as [clinicaltrials.gov](http://clinicaltrials.gov), therefore modern clinical trials lack the transparency needed to support the timely implementation of novel designs. The aim of this manuscript is to bring to light published examples of novel design applications as a means of augmenting the implementation of innovative designs in the future and to demonstrate the flexibility of adaptive designs in satisfying changing design conditions. Displays of current trials that use novel designs are needed to overcome barriers of infrequent implementation of innovative design strategies in early phase trials, so we believe that the current work can aid in the uptake of novel design use. In addition, given the often lengthy timeline between study concepts to protocol completion it is valuable to present design considerations that have a broad application. It is worth noting that even after study completion, journals do not require complete protocols as supplemental material for dose-finding trials, and final clinical trial publications do not have sufficient room to describe the details of novel designs. Therefore, we feel the message that novel methods are being used in clinical practice is a timely and important one.

## 2. Methods

### 2.1. Design considerations

Breast 49 (NCT03473639) is an early-phase evaluation of the safety and tolerability of the combination of two doses (3 mg, 5 mg) of the histone deacetylase inhibitor, entinostat, with two doses of capecitabine (800 mg/m<sup>2</sup>, 1000 mg/m<sup>2</sup>). The trial was initially proposed as a study of the two doses of entinostat in combination with a fixed dose of 1000 mg/m<sup>2</sup> of capecitabine. Discussions with investigators lead to a consensus that the best combination could consist of a lower dose of capecitabine in combination with entinostat, and therefore additional combinations are worth exploring. Fixing the path of escalation based on a pre-specified dose of capecitabine limits the number of combinations that can be considered, potentially missing promising combinations located outside the chosen path. Treatment combinations are grouped into “zones” based on the dose levels of each agent; i.e., zone 1 = {(3 mg, 800 mg/m<sup>2</sup>)}, zone 2 = {(5 mg, 800 mg/m<sup>2</sup>), (3 mg, 1000 mg/m<sup>2</sup>)}, and zone 3 = {(5 mg, 1000 mg/m<sup>2</sup>)} (Table 1). The trial is designed to determine the MTDC, to estimate the proportion of participants able to

tolerate the combination, and to obtain preliminary data on disease-free survival. The study is currently open to accrual and has accrued five participants as of December 18, 2019. A Bayesian adaptive design is being used to guide accrual decisions based on the occurrence of dose-limiting toxicities (DLT’s), and the minimum follow-up period for determination of escalation is 3 weeks. In monitoring safety, adverse events are being assessed and acute toxicity graded using the National Cancer Institute (NCI) Common Terminology Criteria (CTCAE) Version 4.03. A participant is classified as experiencing a DLT (yes/no) based on protocol-defined DLT criteria occurring during the first cycle of treatment. As data accumulates each participant is classified as experiencing a DLT (yes/no), and the MTDC is defined as the dose combination with a DLT rate closest to the target DLT rate of 25%.

The initial dose escalation phase (Part A) of the study will accrue participants in Population A utilizing a model-based design for dose-escalation to determine the initial recommended MTDC. The expansion phase (Part B) will accrue participants in Population B. Expansion will start at the MTDC estimated in Part A. The adaptive modeling strategy will continue to be used in Part B to establish the MTDC in Population B and to estimate the rate of treatment discontinuation at the established MTDC. The initial MTDC (Part A) is proposed when a 7th participant in Population A is recommended to a dose combination that has already accrued 6 participants. Given the limited time window to respond to the FDA’s requested revision, we anticipated that they would prefer a minimum of 6 participants be accrued to the recommended combination in Population A prior to transitioning to Part B. At this point accrual to Part B will begin. The modeling strategy will continue as in Part A to assess safety and establish the MTDC in Population B using accumulated data from both populations. The final MTDC is defined as the combination that is recommended when a 31st participant is recommended to a dose combination that has already accrued 30 participants in Population B. A schema illustrating the proposed trial design is provided in Fig. 1.

### 2.2. Estimation

Safety assessments are based on the assumption that, as the dose level of one agent increases, holding the other agent fixed, the probability of DLT is increasing. It is reasonable to assume that combinations in higher zones have higher probabilities of DLT than combinations in lower zones. It is unknown whether combinations have higher or lower DLT probabilities than other combinations within the same zone. It could be that combination 2 < combination 3 or that combination 3 < combination 2 in terms of their respective true DLT probabilities (see Table 1). We express this uncertainty through specification of two one-parameter models, indexed by  $m$ , that reflect the different ordering possibilities for the DLT probabilities. We then rely on model selection techniques to choose the model most consistent with the data [1]. A common model choice [7] in the continual reassessment method (CRM) is to raise a set of pre-specified constants, also referred to as the ‘skeleton’ of the model, to a power  $\exp(a)$ , where  $a$  is a parameter to be estimated by the data. Denoting the probability of DLT at combination  $i$  as  $R(d_i)$ , we express this class of working models as  $R(d_i) = \Pr(\text{DLT at combination } i) \approx \alpha_{mi}^{\exp(a_m)}$ . For ordering possibility 1, the working model ( $m = 1$ ) is  $(0.25^{\exp(a_1)}, 0.35^{\exp(a_1)}, 0.46^{\exp(a_1)}, 0.56^{\exp(a_1)})$  and for ordering possibility 2, the working model ( $m = 2$ ) is  $(0.25^{\exp(a_2)}, 0.46^{\exp(a_2)}, 0.35^{\exp(a_2)}, 0.56^{\exp(a_2)})$ . The skeleton values for

**Table 1**  
Combination and zone designation.

Entinostat	5 mg	Combination 2	Combination 4
	3 mg	Combination 1	Combination 3
		800 mg/m <sup>2</sup>	1000 mg/m <sup>2</sup>
		<b>Capecitabine</b>	

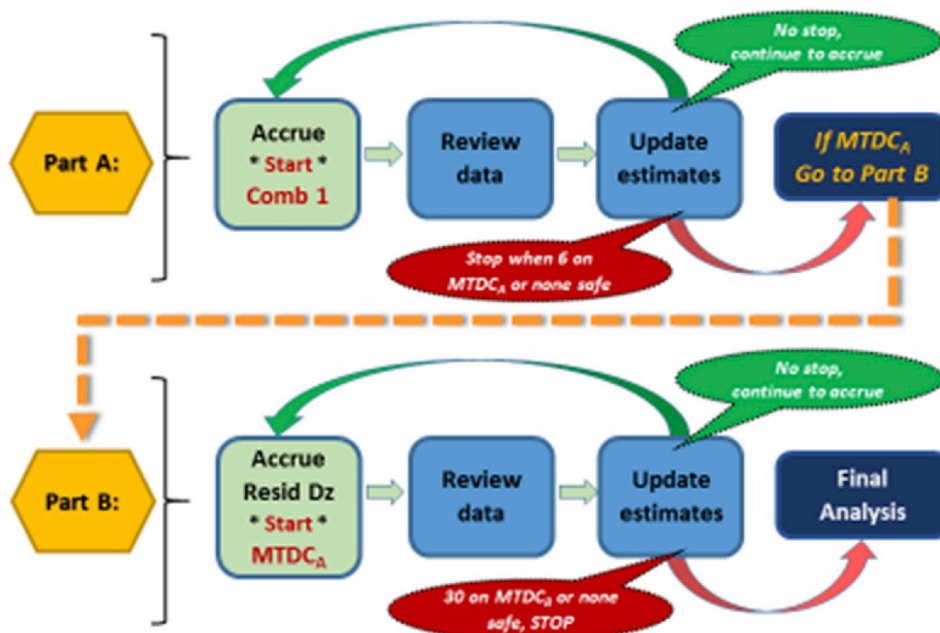


Fig. 1. Schema of illustrating the proposed trial design.

each model were generated using the algorithm of Lee and Cheung [8], and were chosen in order to yield good operating characteristics in a broad range of scenarios. Within each ordering, the CRM is fit using the working model and the accumulated data. This study will accrue eligible participants in cohorts of one and use the fit CRM model to estimate DLT probabilities at each combination. The prior distribution  $g(a_m)$  on the parameters  $a_1$  and  $a_2$  is given by a Normal distribution with mean 0 and variance 1.34; i.e.  $a_m \sim N(0, 1.34)$ , which is common to Bayesian CRM designs and a default distribution utilized in CRM software [9]. A uniform prior distribution  $\tau(m) = 1/m$  is placed on each possible ordering so that each working model is considered equally likely a priori. Based on observed data  $D = \{(y_i, n_i); i = 1, \dots, 4\}$ , where  $y_i$  is the number of DLT's and  $n_i$  is the number of subjects treated on combination  $i$ , the likelihood under working model  $m$  is

$$L_m(D|a_m) \propto \prod_{i=1}^4 (\alpha_{mi}^{\exp(a_m)})^{y_i} (1 - \alpha_{mi}^{\exp(a_m)})^{n_i - y_i}$$

Using Bayes theorem, the posterior probabilities of the models given the data are given by

$$P(m|D) = \frac{\tau(m) \int L_m(D|a_m) g(a_m) da_m}{\sum_{m=1}^2 \tau(m) \int L_m(D|a_m) g(a_m) da_m}$$

After accrual of each participant into the trial, using Bayesian model selection, the working model with the largest posterior model probability  $P(m|D)$  is chosen and the DLT probability estimates  $\hat{R}(d_i)$  are updated using the chosen working model and the Bayesian form of the CRM [10]. If there is a tie between the posterior model probabilities of the two models, then the selected model is chosen at random. Based on the expectedness of events, the target DLT probability is set at 25%. The next participant is allocated to the dose combination indicated by the modeling to have the estimated DLT rate closest to 25%. After each participant, a new recommended combination is obtained, and the next entered participant is allocated to the recommended combination. The trial will stop once sufficient information about the recommended MTDC has been obtained, according to the prespecified stopping rules outlined below.

### 2.3. Stopping the trial

A 90% probability interval is calculated around the estimated DLT

probability for each combination, based on interval estimation for CRM models [11]. If, for the lowest combination, the lower bound of this confidence interval exceeds the target DLT rate of 25%, then combination 1 will be deemed too toxic, the trial will stop for safety, and no combination will be recommended as the MTDC for either study population. Otherwise, accrual to the study will continue until 30 eligible participants in Population B have been treated at the recommended MTDC.

### 2.4. Sample size and accrual

Maximum target sample size is calculated for the goals of (1) establishing the MTDC in participants with residual breast cancer (Population B), (2) obtaining an estimate of treatment tolerance and (3) assessing disease-free survival. Simulation results indicate a maximum target accrual of 55 eligible participants has good properties in terms of accurately identifying the MTDC. Based upon the simulation results, the study goals are achieved with accrual of approximately 12 and 40 (median values; Table 5) participants in populations A and B, respectively. Accrual is estimated at 1–2 participants per month, depending upon which study population is being accrued.

Simulation results indicate that accrual of up to 43 participants in Population B may be required in order to have 30 eligible participants treated at the MTDC. The target of 30 participants is based upon having sufficient information to test for a null rate of treatment tolerance of 60% versus the alternative rate of 80% with a one-sided type I error rate of 0.094 and power of 0.871 with a binomial test. The choice of the null and alternative rates are based upon results reported in the CREATE-X trial [12] which reported 75% (95% CI [69, 80%]) of participants treated with 8 cycles of capecitabine completing treatment. For this study, if the data support a tolerance rate of 60% or lower (lower than the lower limit of the confidence bound), the combination would be considered unacceptable.

## 3. Results

### 3.1. Design behavior early in the trial

In assessing the statistical properties of the design, we first evaluated its behavior early in the trial in order to give investigators and reviewers

**Table 2**  
Sample for when the first two participants in Population A have DLTs.

Participant	Combo	DLT	DLT probability estimates				C.I. lower bound	Recommendation for next participant
			1	2	3	4		
1	1	Yes	0.593	0.746	0.676	0.803	0.12	Combination 1
2	1	Yes	0.690	0.812	0.758	0.856	0.26	STOP STUDY

**Table 3**  
Sample for when the first participant Population A has a DLT, followed by consecutive non-DLTs.

Participant	Combo	DLT	DLT probability estimates				C.I. lower bound	Recommendation for next participant
			1	2	3	4		
1	1	Yes	0.593	0.746	0.676	0.803	0.12	Combination 1
2	1	No	0.449	0.549	0.638	0.714	0.07	Combination 1
3	1	No	0.348	0.453	0.554	0.643	0.05	Combination 1
4	1	No	0.279	0.489	0.384	0.586	0.04	Combination 1
5	1	No	0.230	0.333	0.439	0.541	0.03	Combination 1
6	1	No	0.194	0.294	0.400	0.504	0.03	Combination 2

**Table 4**  
A simulated sequential trial illustrating the described design. The maximum tolerated dose combination (MTDC) recommendation is combination 2 after 53 participants.

Participant	Population	Combo	DLT	Posterior		Participant	Population	Combo	DLT	Posterior	
				prob of ordering						prob of ordering	
				1	2					1	2
1	A	1	no	0.50	0.50	28	B	2	no	0.40	0.60
2	A	2	no	0.53	0.47	29	B	2	yes	0.30	0.70
3	A	3	no	0.50	0.50	30	B	2	yes	0.23	0.77
4	A	4	no	0.50	0.50	31	B	2	yes	0.18	0.82
5	A	4	no	0.50	0.50	32	B	3	no	0.16	0.84
6	A	4	yes	0.50	0.50	33	B	2	no	0.17	0.83
7	A	2	no	0.52	0.48	34	B	2	no	0.19	0.81
8	A	4	no	0.52	0.48	35	B	2	no	0.20	0.80
9	A	4	no	0.52	0.48	36	B	2	no	0.21	0.79
10	A	4	yes	0.53	0.47	37	B	2	yes	0.17	0.83
11	B	4	yes	0.53	0.47	38	B	2	no	0.18	0.82
12	B	3	no	0.50	0.50	39	B	2	no	0.20	0.80
13	B	2	no	0.53	0.47	40	B	2	no	0.21	0.79
14	B	3	no	0.50	0.50	41	B	2	no	0.22	0.78
15	B	3	no	0.48	0.52	42	B	2	yes	0.18	0.82
16	B	4	no	0.48	0.52	43	B	2	yes	0.15	0.85
17	B	4	no	0.48	0.52	44	B	2	no	0.16	0.84
18	B	4	yes	0.47	0.53	45	B	2	no	0.17	0.83
19	B	2	no	0.50	0.50	46	B	2	no	0.18	0.82
20	B	4	yes	0.50	0.50	47	B	2	no	0.19	0.81
21	B	2	no	0.53	0.47	48	B	2	no	0.20	0.80
22	B	3	no	0.50	0.50	49	B	2	no	0.21	0.79
23	B	3	no	0.47	0.53	50	B	2	yes	0.18	0.82
24	B	4	no	0.48	0.52	51	B	2	no	0.19	0.81
25	B	4	yes	0.47	0.53	52	B	2	no	0.20	0.80
26	B	2	yes	0.35	0.65	53	B	2	no	0.21	0.79
27	B	2	no	0.37	0.63						

an idea of how the method begins. In the protocol document, we included two different scenarios illustrating early study behavior. [Tables 2 and 3](#) report the participant index, the combination received by each participant, whether the participant had a DLT, the model-based estimated DLT probabilities, the lower bound of the 90% confidence interval in order to assess whether the trial should stop, and the recommended combination for each participant. As displayed in [Table 2](#), the probability estimates indicate that the trial will terminate for safety if the first two participants experience a DLT on combination 1. [Table 3](#) illustrates the results for a different situation in which the first participant experiences a DLT and we want to see how many consecutive non-DLTs on combination 1, starting with the second participant, would be required to escalate to either combination 2 or 3. If the first participant has a DLT on combination 1, we would be required to observe five

consecutive non-DLTs on combination 1 prior to being allowed to escalate to combination 2.

### 3.2. Simulated trial example

In this section, we illustrate the behavior of the design described in this article under a set of hypothesized DLT probabilities. The assumed DLT probabilities for combinations 1–4 in Population A are (0.14, 0.35, 0.22, 0.50) and the assumed DLT probabilities in Population B are (0.04, 0.25, 0.12, 0.40), indicating combination 2 to be the true MTDC in Population B since it has the DLT probability closest to the target rate of 25%. The true underlying DLT probabilities are consistent with ordering 2 in which combination 2 is more toxic than combination 3. The data from the entire simulated trial are provided in [Table 4](#). The first eligible

**Table 5**

Simulation studies of design operating characteristics. The results displayed are based upon a maximum target accrual of 55 participants.

Scenario	Entinostat	True DLT probability (Pop A, Pop B)	% recommended	% Stop	% DLT	Sample size Met	Pts 25th	Sample size Resid	Pts 25th	Total sample size	25th %	
		800 mg/m <sup>2</sup>	1000 mg/m <sup>2</sup>			50th %		50th %		50th %		
		Capecitabine				75th %		75th %		75th %		
1 All safe	5 mg	(0.09,0.09)	(0.15, <b>0.15</b> )	0.1	13.1	9		30		39		
		0.8	<b>98.5</b>			10		30		42		
		4.1	<b>34.0</b>			14		32		46		
	3 mg	(0.02,0.02)	(0.09,0.09)									
		0.0	0.6									
		1.7	3.7									
2 All but one safe	5 mg	(0.15,0.15)	(0.35,0.35)	0.9	22.4	9		35		49		
		25.5	22.8			12		40		55		
		14.4	12.2			15		43		55		
	3 mg	(0.10,0.10)	(0.25, <b>0.25</b> )									
		2.5	<b>48.3</b>									
		5.6	<b>19.2</b>									
3 All but one safe	5 mg	(0.25, <b>0.25</b> )	(0.35,0.35)	1.2	22.8	10		36		49		
		47.6	23.5			12		40		55		
		19.3	11.8			15		43		55		
	3 mg	(0.10,0.10)	(0.15,0.15)									
		1.1	26.6									
		4.7	15.5									
4 Only one safe	5 mg	(0.50,0.50)	(0.60,0.60)	8.5	28.6	7		31		41		
		5.0	0.0			10		38		49		
		5.2	0.6			13		42		55		
	3 mg	(0.20, <b>0.20</b> )	(0.37,0.37)									
		66.8	19.7									
		26.3	13.4									
5 All toxic	5 mg	(0.60,0.60)	(0.70,0.70)	96.7	51.7	2		0		2		
		0.0	0.0			6		2		8.5		
		0.4	0.1			8		8		16		
	3 mg	(0.50,0.50)	(0.60,0.60)									
		3.3	0.0									
		10.1	1.0									
6 All but one safe in Pop B	5 mg	(0.50,0.15)	(0.60,0.35)	10.1	23.0	7		40		55		
		31.1	6.6			9		44		55		
		12.7	2.8			13		47		55		
	3 mg	(0.20,0.10)	(0.37, <b>0.25</b> )									
		12.1	<b>47.2</b>									
		18.8	<b>17.7</b>									
7 All safe in Pop B	5 mg	(0.25,0.09)	(0.35, <b>0.15</b> )	1.8	15.7	9		38		55		
		13.1	<b>79.3</b>			13		41		55		
		13.2	<b>20.0</b>			14		43		55		
	3 mg	(0.10,0.02)	(0.15,0.09)									
		0.0	7.3									
		6.9	12.7									

Pop A = Population A; Pop B = Population B; avg = average; pts = participants.  
 % Stop = percentage of trials stopped early for safety with no MTDC recommendation.  
 Met = metastatic disease; Resid = residual disease.  
 25th, 50th, 75th % = 25th, 50th, 75th percentiles.

participant in Population A is allocated to combination 1 and escalation proceeds without DLT until participant 6 in Population A experiences a DLT on combination 4. The design then recommends de-escalating to combination 2 on which an additional non-DLT is observed. The design recommends returning to combination 4 on which DLT outcomes are observed in 1 out of the next 3 participants. In total, 6 participants in Population A have now been treated at combination 4 and this combination is recommended for the next participant. At this point, accrual to Population A is terminated with combination 4 being defined as the starting combination for Population B. In Population A, the observed DLT data at combinations 1–4, respectively, are (0/1, 0/2, 0/1, 2/6). Without much data accumulated at combinations 2 and 3 at this point, the posterior model probabilities associated with orderings 1 and 2,

respectively, are (0.53, 0.47). Based on this data, ordering 1 is selected as the DLT probability ordering most consistent with the data, and the posterior mean of  $a_1$  is 0.73. The model-based estimated DLT probabilities are  $(0.25^{\exp(0.73)}, 0.46^{\exp(0.73)}, 0.35^{\exp(0.73)}, 0.56^{\exp(0.73)}) = (0.056, 0.113, 0.199, 0.300)$ , indicating that combination 4 has DLT probability closest to the target rate of 25%. In accruing participants in Population B, several DLTs are observed at combination 4 (participants 11, 18, 20, and 25), prompting the design to settle in Zone 2 and to accumulate data on combinations 2 and 3. In the presence of DLTs observed on combination 2 (participants 26, 29, 30, and 31), the posterior model probabilities begin to separate from one another (0.18, 0.82), indicating ordering 2 to be more likely to represent the true DLT probability ordering. Ultimately, the design settles on combination 2 for

the final 21 participants accrued to the study, triggering the stopping rule once 30 participants in Population B are accrued to combination 2, and recommending combination 2 as the MTDC. At study conclusion, the total observed DLT data from both populations at combinations 1–4, respectively, are (0/1, 8/32, 0/7, 6/13) with model-based estimated DLT probabilities (0.087, **0.254**, 0.157, 0.359).

### 3.3. Operating characteristics

In addition to the illustrations above, we conducted computer simulation studies in order to evaluate the operating characteristics of the design described in the previous section over a broad range of assumed combination-toxicity scenarios with the following “themes,” (1) all assumed DLT probabilities are acceptable in terms of safety (i.e.  $\leq 25\%$ ), (2) all combinations but one are acceptable in terms of safety, (3) only one combination is acceptable in terms of safety, and (4) all combinations are too toxic. For each scenario, 5000 simulated trials were run. Table 5 reports the true DLT probability at each combination, the percentage of trials in which each combination was recommended as the MTDC, and the average number of participants treated at each combination. Displayed in the last five columns are the percentage of times in the simulations that the trial closed due to safety concerns, the percentage of simulated participants that had a DLT, and the selected percentiles for the number of participants for each participant population and total trial size at study closure. The results displayed in Table 5 were based upon a maximum target accrual of 55 participants, and accrual to the study was stopped when 30 eligible participants in Population B were treated at the recommended MTDC. With this type of design and stopping rules, the results indicated that on average the trial would achieve this goal with accrual in the range of approximately 42–55 participants.

From examining the results in Table 5 the proposed design has good properties in terms of recommending optimal dose combinations, as well as allocating participants to these combinations. In Scenario 1, the design selects the true MTDC in 98.5% of simulated trials, while assigning 34.0 participants on average to this combination with a median trial size of 42 participants. In Scenario 2, recommendation of the true MTDC occurs in approximately 48.3% of simulated trials based on a median trial size of 55 participants, while allocating 19.2 participants on average to the true MTDC. It is important to note that when the target combination is not selected as the MTDC, treatments with assumed DLT probabilities within an acceptable toxicity range of 15%–35% are selected in 96.6% of simulated trials. Similar findings are obtained for Scenario 3. In Scenario 4, the design identifies the target combination as the MTDC in approximately 66.8% of simulated trials based on a median total trial size of 49, while allocating 26.3 participants on average to this combination. When combination 1 is not selected, the method tends to either choose combination 2 with the next highest assumed DLT rate (19.7% of the time), or stops the trial for safety (8.5% of the time). In Scenario 5, where all combinations are overly toxic, the method correctly terminates the study in 96.7% of simulated trials based on a maximum trial size of 16 participants, and treats 10.1 accrued participants on average to Zone 1. Scenarios 6 and 7 present cases in which the DLT probabilities differ between the two populations, under the assumption that Population A has higher DLT probabilities than Population B. In Scenario 6, recommendation of the true MTDC occurs in approximately 47.2% of simulated trials based on a median trial size of 55 participants, while allocating 17.7 participants on average to the true MTDC. Finally, in Scenario 7, the design selects the true MTDC in 79.3% of simulated trials, while assigning 20.0 participants on average to this combination with a median trial size of 55 participants. The performance of the design in Scenario 6 and 7 diminishes slightly when compared with Scenarios 2 and 1, respectively, as expected. Scenarios 6 and 7 have the same Population B DLT probabilities as Scenarios 2 and 1, but have different Population A probabilities that place the true MTDC in Population A at a different combination than its location in

Population B. Despite differing probabilities and location of MTDCs between populations, the design is still able to correctly identify the MTDC at the conclusion of the study in a high percentage of simulated trials. Overall, the simulation results indicate that the design outlined in this article is a practical early-phase adaptive method for use with drug combination therapies in the presence of heterogeneous patient populations.

## 4. Conclusions

The growth of novel methods in early-phase dose-finding has been rapid in the last decade, yet the implementation of innovative designs remains uncommon. In this article, we have outlined a novel adaptive design for early-phase trials involving a shift in patient populations implemented in an ongoing trial of a novel drug combination for participants with metastatic breast cancer and residual breast cancer after neo-adjuvant chemotherapy. The model-based method presented serves as an alternative to rigid rule-based methods that lack the ability to handle the complexity presented both by drug combination trials, as well as by patient heterogeneity. The use of more innovative approaches are being encouraged by the FDA and by others [13–16]. Simulation studies were performed to justify and evaluate the performance of the design characteristics. The simulation results in Table 5 demonstrate the method’s ability to effectively recommend desirable combinations, defined by acceptable toxicity, in a high percentage of trials with manageable sample sizes. We also explored the possibility of only using the data from Population A in order to arrive at the starting dose combination for Population B and ignoring the Population A data for estimation in the expansion phase. Simulation studies over a broad range of assumed combination-toxicity scenarios (results not shown) demonstrated that ignoring the Population A data for estimation in the expansion phase yielded slightly better performance in terms of correctly identifying the true MTDC, but at the expense of yielding an overall lower probability of accruing 30 participants on a combination in Population B. With the aim of accumulating enough information to evaluate treatment tolerance and disease-free survival at the selected MTDC, we wanted the design to appropriately balance the objectives of correctly identifying the MTDC and accruing 30 participants to this dose combination. Our assessment of these two approaches concluded that continuing to use the Population A data for estimation in the expansion phase was a more efficient use of the accumulating data.

Software in the form of R [17] code for both simulation of design operating characteristics and direct protocol implementation of the method is available at [http://faculty.virginia.edu/model-based\\_dose-finding/](http://faculty.virginia.edu/model-based_dose-finding/). The method we outline in this work can be viewed as an extension of the CRM, utilizing multiple skeletons for DLT probabilities to account for the uncertainty surrounding the toxicity profile of drug combinations. This increases the flexibility of CRM designs, enabling it to handle more complex dose-finding problems. The numerical results presented in the simulation studies such as the distribution of sample size and frequency of early trial termination is the type of simulation information that improve understanding, acceptance, and approval of novel designs such as the one described in this manuscript [18,19]. This support for adaptive designs will augment efficient early-phase trial design in drug combination studies [20]. Well-performing dose-finding designs can have a tremendous impact on the drug development process [21].

## Financial support

This work is supported by the National Cancer Institute [K25CA181638 to N.A.W. and R01CA142859 to G.R.P.]; the Biostatistics Shared Resource, University of Virginia Cancer Center, University of Virginia [P30 CA044579].

## Declaration of competing interest

The authors declare no potential conflicts of interest.

## References

- [1] N.A. Wages, M.R. Conaway, J. O'Quigley, Continual reassessment method for partial ordering, *Biometrics* 67 (2011) 1555–1563.
- [2] N.A. Wages, C.L. Slingluff Jr., G.R. Petroni, A Phase I/II adaptive design to determine the optimal treatment regimen from a set of combination immunotherapies in high-risk melanoma, *Contemp. Clin. Trials* 41 (2015) 172–179.
- [3] N.A. Wages, C.A. Portell, M.E. Williams, M.R. Conaway, G.R. Petroni, Implementation of a model-based design in a phase Ib study of combined targeted agents, *Clin. Cancer Res.* 23 (2017) 7158–7164.
- [4] N.A. Wages, C.L. Slingluff Jr., G.R. Petroni, Statistical controversies in clinical research: early-phase adaptive design for combination immunotherapies, *Ann. Oncol.* 28 (2017) 696–701.
- [5] A. Iasonos, J. O'Quigley, Design considerations for dose expansion cohorts in phase I trials, *J. Clin. Oncol.* 31 (2013) 4014–4021.
- [6] P.S. Boonstra, J. Shen, J.M.G. Taylor, et al., A statistical evaluation of dose expansion cohorts in phase I clinical trials, *J. Natl. Cancer Inst.* 107 (2015) dju429.
- [7] X. Paoletti, A. Kramar, A comparison of model choices for the continual reassessment method in phase I clinical trials, *Stat. Med.* 28 (2009) 3012–3028.
- [8] S.M. Lee, Y.K. Cheung, Model calibration in the continual reassessment method, *Clin. Trials* 6 (2009) 227–238.
- [9] Y.K. Cheung, *Dose Finding by the Continual Reassessment Method*, CRC Press, New York, New York, 2011.
- [10] J. O'Quigley, M. Pepe, L. Fisher, Continual reassessment method: a practical design for phase I clinical trials in cancer, *Biometrics* 46 (1990) 33–48.
- [11] L. Natarajan, J. O'Quigley, Interval estimates of the probability of toxicity at the maximum tolerated dose for small samples, *Stat. Med.* 22 (2003) 1829–1836.
- [12] N. Masuda, S.-J. Lee, S. Ohtani, et al., Adjuvant capecitabine for breast cancer after preoperative chemotherapy, *N. Engl. J. Med.* 376 (2017) 2147–2159.
- [13] M.K. Riviere, C. Le Tourneau, X. Paoletti, et al., Designs of drug-combination phase I trials in oncology: a systematic review of the literature, *Ann. Oncol.* 26 (2015) 1036–1037.
- [14] X. Paoletti, M. Ezzalfani, C. Le Tourneau, Statistical controversies in clinical research: requiem for the 3 + 3 design for phase I trials, *Ann. Oncol.* 26 (2015) 1808–1812.
- [15] A. Iasonos, J. O'Quigley, Adaptive dose-finding studies: a review of model-guided phase I clinical trials, *J. Clin. Oncol.* 32 (2014) 2505–2511.
- [16] L. Nie, E.H. Rubin, N. Mehrotra, et al., Rendering the 3+3 design to rest: more efficient approaches to oncology dose-finding trials in the era of targeted therapy, *Clin. Cancer Res.* 22 (2016) 2623–2629.
- [17] R Development Core Team, *R: A Language and Environment for Statistical Computing*, R Foundation for Statistical Computing, Vienna, Austria, 2008.
- [18] A. Iasonos, M. Gönen, G.J. Bosl, Scientific review of phase I protocols with novel dose-escalation designs: how much information is needed? *J. Clin. Oncol.* 33 (2015) 2221–2225.
- [19] G.R. Petroni, N.A. Wages, G. Paux, et al., Implementation of adaptive methods in early-phase clinical trials, *Stat. Med.* 36 (2017) 215–224.
- [20] N.A. Wages, M.R. Conaway, C.L. Slingluff Jr., et al., Recent developments in the implementation of novel designs for early-phase combination studies, *Ann. Oncol.* 26 (2015) 1036–1037.
- [21] M.R. Conaway, G.R. Petroni, The impact of early phase trial design in the drug development process, *Clin. Cancer Res.* 25 (2019) 819–827.